Business Analytics (110-1)

Assignment 4

Due: 9:00 am, Tue 21-Dec-2021

1.

For the data in "Smoking & Lung Cancer" case, test whether the odds of lung cancer for smokers are equal to the odds of lung cancer for nonsmokers, using Fisher's Exact Test.

H0: phi = 1

H1: phi > 1

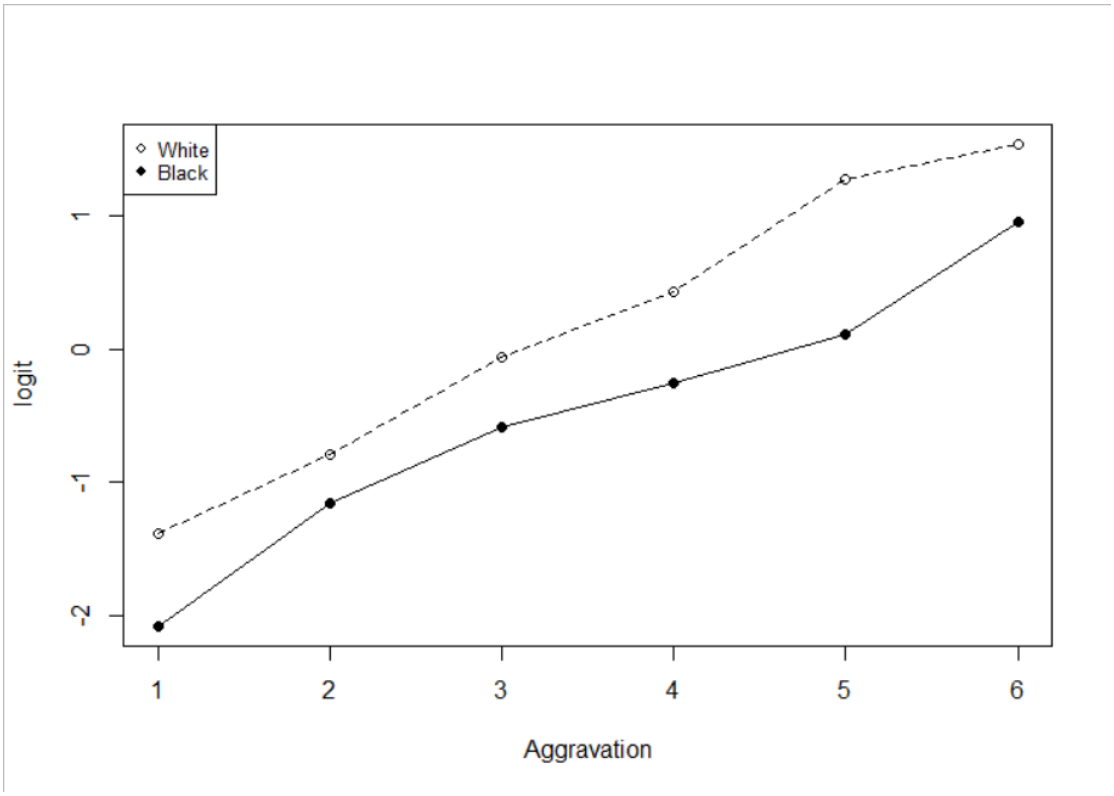
```
> fisher.test(case1803, alternative = 'greater')

        Fisher's Exact Test for Count Data

data:  case1803
p-value = 0.004411
alternative hypothesis: true odds ratio is greater than 1
95 percent confidence interval:
 1.666128      Inf
sample estimates:
odds ratio
  5.333256
```

觀察圖中fisher's exact test結果可知p-value = 0.004411 < 0.05，我們可以拒絕H0，得知odds ratio != 1，兩組的odds並不相等。

2.

Reconsider the case Death Penalty and Race of Murder Victim discussed in lecture 10 (case1902). Reanalyze the data **using logistic regression**. The response variable is the number of convicted murderers in each category who receive the death sentence, out of the **m convicted murderers** in that category.

(a) Plot the logits of the observed proportions versus the level of aggravation. The logit, however, is undefined for the rows where the proportion is 0 or 1, so **compute the empirical logit = log[(y + 0.5) / (m − y + 0.5)] and plot this versus aggravation level**, using different plotting symbols to distinguish proportions based on white and black victims.



(b) Fit **the logistic regression of death sentence proportions** on aggravation level and an indicator variable for race of victim.

```
Call:
glm(formula = cbind(Death, Nodeath) ~ Victim + Aggravation, family = binomial,
    data = case1902_new)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-0.93570  -0.22548   0.05142   0.65620   1.01444

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -4.8653     0.6004  -8.103 5.37e-16 ***
VictimBlack  -1.8106     0.5361  -3.377 0.000732 ***
Aggravation   1.5397     0.1867   8.246  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 212.2838  on 11  degrees of freedom
Residual deviance:   3.8816  on  9  degrees of freedom
AIC: 31.747

Number of Fisher Scoring iterations: 4
```

(c) Report the p-value from the deviance goodness-of-fit test for this fit.

H0: the chosen Logistic GLM holds

```
> # 2-c
> deviance(model)
[1] 3.881585
> model$df.residual
[1] 9
> 1 - pchisq(deviance(model), model$df.residual)
[1] 0.9190319
```

p-value = 0.919 > 0.05, there is no evidence that the model is inadequate

(d) Test whether the coefficient of the indicator variable for race is equal to 0, using Wald's test.

H0: Beta = 0

H1: Beta != 0

```
Call:
glm(formula = cbind(Death, Nodeath) ~ Victim + Aggravation, family = binomial,
    data = case1902_new)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-0.93570  -0.22548   0.05142   0.65620   1.01444

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -4.8653     0.6004  -8.103 5.37e-16 ***
VictimBlack  -1.8106     0.5361  -3.377 0.000732 ***
Aggravation   1.5397     0.1867   8.246  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 212.2838  on 11  degrees of freedom
Residual deviance:   3.8816  on  9  degrees of freedom
AIC: 31.747

Number of Fisher Scoring iterations: 4
```

VictimBlack的p-value為0.000732 < 0.05，因此可以推翻H0，coefficient of the indicator variable for race is not equal to 0.

(e) Construct a confidence interval for the same coefficient, and **interpret it in a sentence about the odds of death sentence for white-victim murderers relative to black-victim murderers**, accounting for aggravation level of crime.

```
                 2.5 %     97.5 %
(Intercept) -6.171218 -3.7937756
VictimBlack -2.929963 -0.8021809
Aggravation  1.207083  1.9464923
```

(f) Refit the model by treating the aggravation level as a factor. How would you interpret the results of this model?

```
Call:
glm(formula = cbind(Death, Nodeath) ~ Victim + as.factor(Aggravation),
    family = binomial, data = case1902_new)

Deviance Residuals:
       1         2         3         4         5         6         7         8         9
 0.02705  -0.03705  -0.27695   0.46062  -0.22255   0.33222   0.02846  -0.03695   1.21437
      10        11        12
-0.55797   0.00006   0.00007

Coefficients:
                         Estimate Std. Error z value Pr(>|z|)
(Intercept)               -3.4207     0.6144  -5.567 2.59e-08 ***
VictimBlack               -1.7409     0.5426  -3.208  0.00134 **
as.factor(Aggravation)2    1.6090     0.8506   1.892  0.05855 .
as.factor(Aggravation)3    3.3902     0.7474   4.536 5.74e-06 ***
as.factor(Aggravation)4    4.5004     0.7858   5.727 1.02e-08 ***
as.factor(Aggravation)5    5.8814     0.9128   6.443 1.17e-10 ***
as.factor(Aggravation)6   26.2636  8772.8073   0.003  0.99761
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 212.2838  on 11  degrees of freedom
Residual deviance:   2.2391  on  5  degrees of freedom
AIC: 38.105

Number of Fisher Scoring iterations: 19
```